

WHAT IS CLAIMED IS:

1. A method for speech processing in a code excitation linear prediction (CELP) based speech system having a plurality of modes including at least a first mode and a second mode consecutive with the first mode, comprising:

providing an input speech signal;

dividing the speech signal into a plurality of frames;

dividing at least one of the plurality of frames into sub-frames including a plurality of pulses;

selecting a first number of pulses for the first mode, with a second number of remaining pulses in the frame plus the first number of pulses in the first mode for the second mode;

providing a plurality of sub-modes between the first mode and the second mode, wherein each sub-mode contains a third number of pulses including at least all the pulses in the first mode, and wherein the third number of pulses in the sub-mode are selected by dropping a portion of the pulses in the second mode;

forming a base layer including the first number of pulses;

forming an enhancement layer including the second number of the remaining pulses;

generating a bit stream including a basic bit stream and an enhancement bit stream, including

generating linear prediction coding (LPC) coefficients,

generating pitch-related information,

generating pulse-related information,

forming the basic bit stream including the LPC coefficients, the pitch-related information, and the pulse-related information of the pulses in the base layer, and

forming the enhancement bit stream including the pulse-related information of the pulses in the enhancement layer,

wherein the basic bit stream is used to update memory states of the speech system.

2. The method as claimed in claim 1, wherein the LPC coefficients and the pitch-related information are used to update memory states of the speech system.

3. The method as claimed in claim 1, wherein the pulse-related information of the pulses in the base layer is used to update memory states of the speech system.

4. The method as claimed in claim 1, wherein
generating pulse-related information is based on a fixed codebook, and
generating pitch-related information is based on an adaptive codebook,
wherein the adaptive codebook only contains the information in the basic bit stream.

5. The method as claimed in claim 1, wherein both generating the pitch-related information and generating the pulse-related information comprise minimizing a difference between a synthesized speech and a target signal.

6. The method as claimed in claim 5, wherein the step of minimizing the difference between the synthesized speech and the target signal is looped once for the pulses in each frame to generate the pitch-related information and the pulse-related information for the second number of pulses in the second mode, the first number of pulses from the second mode to form the first mode, and the third number of pulses from the second mode to form the sub-modes.

7. The method as claimed in claim 6, wherein the third number of pulses of each sub-mode are selected by dropping one or more pulses from the second number of pulses in the second mode without the minimization step.

8. The method as claimed in claim 6, wherein the first number of pulses in the first mode are selected by dropping one or more pulses from the third number of pulses of each sub-mode without the minimization step.

9. The method as claimed in claim 1, wherein each sub-mode between the first mode and the second mode corresponds to a second bit stream, wherein the second bit stream is formed by including the basic bit stream and selecting a portion of the enhancement bit stream.

10. The method as claimed in claim 9, wherein the second bit stream includes the pulse-related information of the third number of pulses of each sub-mode, wherein the third number depends on available channel bandwidth.

11. The method as claimed in claim 1, wherein the plurality of sub-modes include at least a first sub-mode and a second sub-mode, wherein the third number of pulses of the first sub-mode are selected by dropping one or more pulses from the second number of pulses in the second mode, and the third number of pulses of the second sub-mode are selected by dropping one or more pulses from the third number of pulses of the first sub-mode.

12. The method as claimed in claim 10, wherein all of the third number of pulses participate in generating a synthesized speech.

13. The method as claimed in claim 11, wherein the pulse dropped between the second mode and the first sub-mode and between consecutive sub-modes are from alternating sub-frames.

14. The method as claimed in claim 13, wherein the pulses dropped from the second mode to constitute the third number of pulses of the first sub-mode are from the first sub-frame, and the pulses dropped from the first sub-mode to

constitute the third number of pulses of the second sub-mode are from the third sub-frame.

15. The method as claimed in claim 11, wherein the dropped pulses are used to transmit non-voice data.

16. A method for transmitting non-voice data together with voice data over a voice channel having a fixed bit rate, comprising:

providing an amount of non-voice data;

providing a speech signal to be transmitted over the voice channel;

dividing the speech signal into a plurality of frames;

dividing at least one of the plurality of frames into sub-frames including a plurality of pulses;

selecting a first number of pulses for the first mode, with a second number of pulses remaining in the frame plus the first number of pulses in the first mode for the second mode;

providing a plurality of sub-modes between the first mode and the second mode, wherein each sub-mode contains a third number of pulses including at least all the pulses in the first mode, and wherein the third number of pulses in each sub-mode are selected by dropping a portion of the pulses in the second mode;

forming a base layer including the first number of pulses;

forming an enhancement layer including the second number of pulses;

forming a first bit stream including a basic bit stream and an enhancement bit stream, including

generating linear prediction coding (LPC) coefficients,

generating pitch-related information,

generating pulse-related information for all of the second number of pulses,

forming the basic bit stream including the LPC coefficients, the pitch-related information, and the pulse-related information of each pulse in the base layer,

selecting one of the sub-modes, and

forming the enhancement bit stream including the pulse-related information of the pulses in selected sub-mode;

forming a second bit stream with the fixed bit rate by including the first bit stream and the amount of the non-voice data; and

transmitting the second bit stream.

17. The method as claimed in claim 16, wherein the voice channel is a channel in an AMR-WB system, the first mode and the second mode are standard modes of the AMR-WB system.

18. The method as claimed in claim 17, wherein all of the first bit stream of the selected sub-mode is used to update memory states of an AMR-WB system.

19. The method as claimed in claim 16, wherein the second bit stream of each sub-mode includes the pulse-related information of a third number of pulses, and the third number of pulses include all of the first number of pulses and are selected by dropping a fourth number of pulses from the second number of pulses.

20. The method as claimed in claim 18, further comprising:
providing an amount of non-voice data; and
modulating the fourth number of dropped pulses of the selected sub-mode with the non-voice data,
transmitting the modulated fourth number of dropped pulses.

21. The method as claimed in claim 18, wherein the third number of pulses of a first sub-mode are selected by dropping one or more pulses from the second mode, and the third number of pulses of a subsequent sub-mode are selected by dropping one or more pulses from a previous sub-mode.

22. The method as claimed in claim 21, wherein the dropped pulses between the first mode and the first sub-mode and between consecutive sub-modes are from alternating sub-frames.

23. The method as claimed in claim 21, wherein the pulses dropped from the second mode to constitute the third number of pulses of the first sub-mode are from the first sub-frame, and the pulses dropped from the first sub-mode to

constitute the third number of pulses of a second sub-mode are from the third sub-frame, etc.